**Phần 2\_3\_Lớp chuẩn hóa**

**Notes:**

* *Về format cấu trúc soạn như sau:*
  + *Lý thuyết…*
  + *Bộ code mấu/ ví dụ …*
  + *Ứng dụng (nếu có)...*
* *Mems làm nhớ note tên để mn dễ contact*

**Mục lục**

[**I. Nội dung chính**](#_geyqu1b9n8lp) **1**

[**II. Nội dung biên soạn chi tiết**](#_sp7nvqu93sa9) **1**

[1. Chuẩn hóa là gì và tại sao nó hữu ích?](#_lo8t59wbwszt) 1

[2. Batch Normalization (BN)](#_rfdlah2jp0x7) 2

[3. Layer Normalization (LN)](#_3v1w964a2kx9) 3

[4. Instance Normalization](#_fn6hs5pcrc6k) 5

### **I. Nội dung chính**

Về các phép tính toán (lớp tính toán) để xây dựng mô hình học sâu.

* + (3) Lớp chuẩn hóa: **batch-norm**, **layer-norm**, **instance-norm**
    - Nguyên lý làm việc
    - Số tham số có thể học
    - Độ rộng của vùng nhận thức (receptive fields)
    - Hình dạng kích thước của bản đồ đặc trưng đầu vào và đầu ra

### **II. Nội dung biên soạn chi tiết**

#### **1. Chuẩn hóa là gì và tại sao nó hữu ích?**

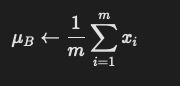
Chuẩn hóa một tập hợp dữ liệu sẽ biến đổi tập hợp dữ liệu thành một tỷ lệ tương tự. Đối với các mô hình học máy, mục tiêu của chúng tôi thường là lấy lại và thay đổi quy mô dữ liệu của chúng tôi sao cho nằm trong khoảng từ 0 đến 1 hoặc -1 đến 1, tùy thuộc vào chính dữ liệu đó. Một cách phổ biến để thực hiện điều này là tính giá trị trung bình và độ lệch chuẩn trên tập hợp dữ liệu và biến đổi từng mẫu bằng cách trừ giá trị trung bình và chia cho độ lệch chuẩn, điều này là tốt nếu chúng ta giả định rằng dữ liệu tuân theo phân phối chuẩn như sau Phương pháp này giúp chúng ta chuẩn hóa dữ liệu và đạt được phân phối chuẩn chuẩn.

Quá trình chuẩn hóa có thể giúp đào tạo mạng lưới thần kinh của chúng ta vì các tính năng khác nhau có quy mô tương tự nhau, giúp ổn định bước giảm độ dốc, cho phép chúng ta sử dụng tốc độ học lớn hơn hoặc giúp các mô hình hội tụ nhanh hơn với tốc độ học nhất định.

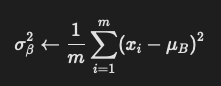
#### **2. Batch Normalization (BN)**

đề cập đến việc chuẩn hóa giá trị input của layer bất kỳ. Chuẩn hóa có nghĩa là đưa phân phối của layer về xấp xỉ phân phối chuẩn với trung bình xấp xỉ 0 và phương sai xấp xỉ 1. Về mặc toán học, Batch Normalization (BN) thực hiện như sau: với mỗi layer, BN tính giá trị trung bình và phương sai của nó. Sau đó sẽ lấy giá trị đặc trưng trừ giá trị trung bình , sau đó chia cho độ lệch chuẩn. Thực tế, chúng ta hay chia tập train thành từng batch với kích thước là 16,32,64 ,128 … hình, hay còn gọi là 1 mini-batch size 16,32,64,128 …. BN được tính toán trên các mini-batch đó.

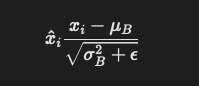
Công thức tính trung bình của mini-batch



Công thức tính phương sai của mini-batch:



Chuẩn hóa:



Phía trên mà mô tả toán học phép biến đổi Batch Normalizing , sử dụng cho hàm kích hoạt x trên mini-batch.

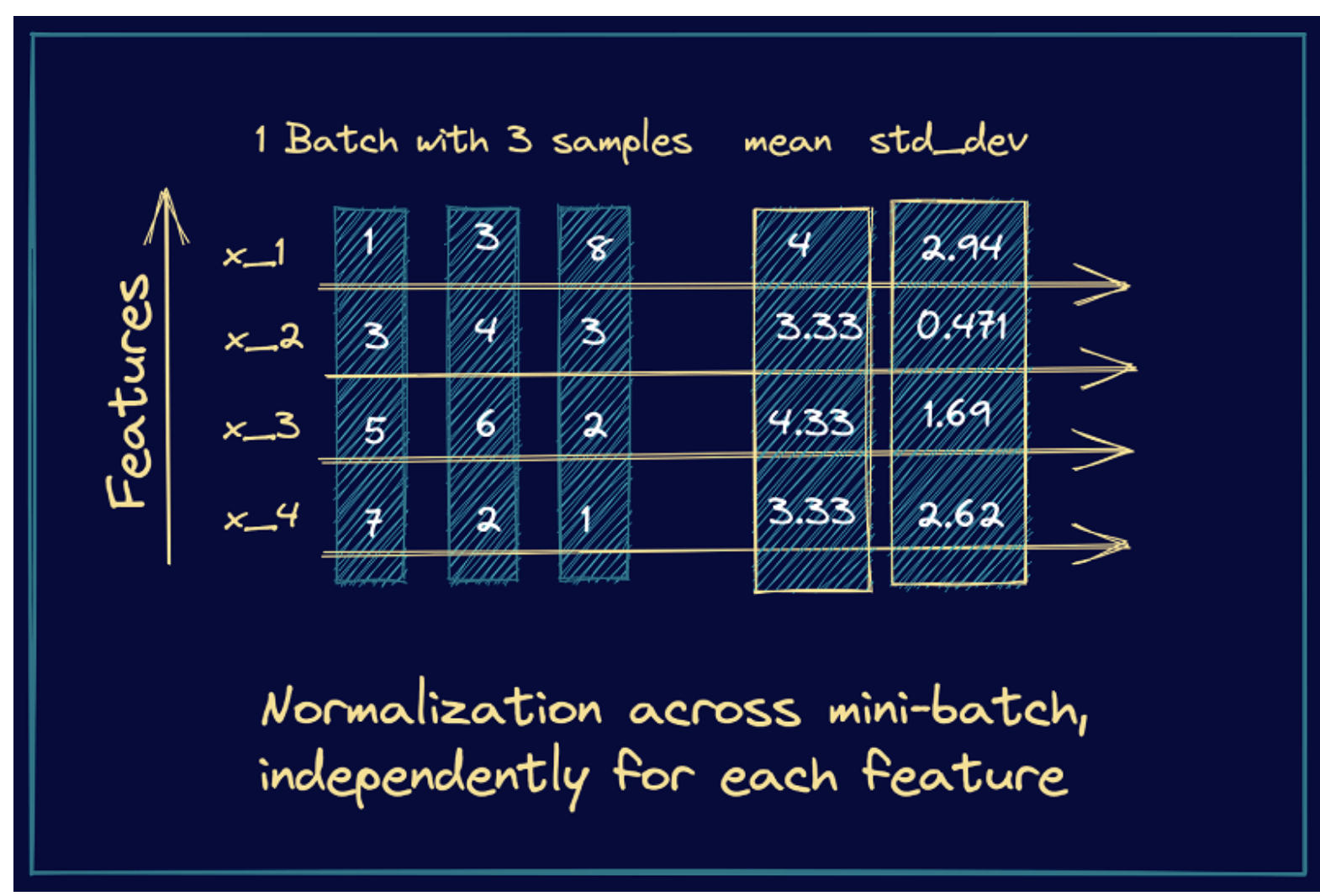
Thực tế, đôi khi mô hình lại hoạt động hiệu quả với một giá trị trung bình và phương sai khác, nên tác giả thêm 2 siêu tham số là gamma - scale và beta - shift để có tính tổng quát.



Batch Normalization hoạt động như thế nào

Về mặc trực quan, chúng ta biết rằng, trong gradient descent, mạng CNN tính giá trị đạo hàm và giảm trọng số của nó dựa vào hướng đi của đạo hàm. Nhưng do các layer được xếp chồng lên nhau, phân phối của dữ liệu đầu vào sẽ bị thay đổi dần do việc cập nhật trọng số của các layer trước đó, làm cho phân phối của đầu vào của các layer phía sau sẽ khác xa so với phân phối của data input. BN giúp cố định phân phối của dữ liệu về phân phối chuẩn, qua tất cả các lớp, dẫn tới tính chất phân phối của dữ liệu không thay đổi qua các lớp.

ví dụ:



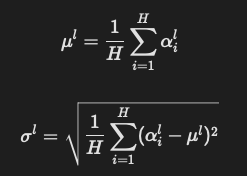
x\_1

mean = (1+3+8)/3 = 4

std\_dev =sqrt(((1-4)^2 + (3-4)^2 + (8-4)^2)/3 ) = 2.94

#### **3. Layer Normalization (LN)**

Phép chuẩn hóa được sử dụng trên từng layer như sau:

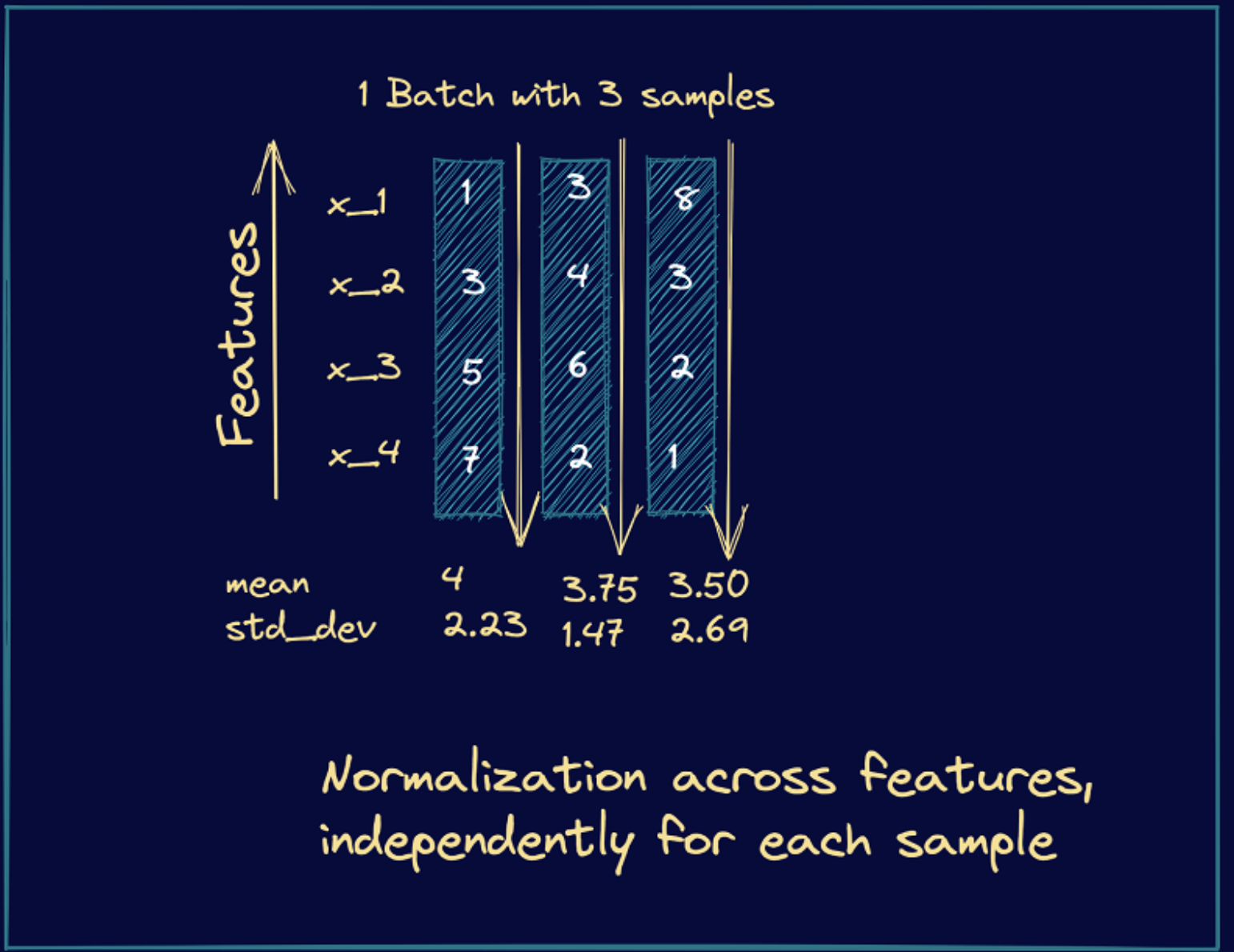


Với H là số lượng phần tử trong một hidden layer.

Cái khác nhau chính giữa BN và LN là LN sử dụng chung một giá trị trung bình và phương sai trong 1 hidden layer. LN không phụ thuộc vào mini-batch, nên có thể train được với batch-size = 1 mà không gặp vấn đề gì cả.

Ngoài ra LN cũng có thể được sử dụng trong RNN mà không gặp trở ngại nào như BN.

ví dụ



(1+3+5+7)/4 = 4

sqrt(((1-4)^2+(3-4)^2+(5-4)^2+(7-4)^2)/4) = 2.23

#### **4. Instance Normalization**

Instance Normalization còn có tên gọi khác là contrast normalization

Ý tưởng ở đây là chúng ta sẽ chuẩn hoá trên từng channel của từng batch.

